

Mixture of Kernels and Iterated Semi-Direct Product of Diffeomorphisms Groups

Martins Bruveris, Laurent Risser and François-Xavier Vialard

August 12, 2011

Contents

1	Introduction	1
2	A finite number of scales	3
2.1	The finite mixture of kernels	3
2.2	Iterated semi-direct product	5
2.3	The Order Reversed	8
3	Extension to a continuum of scales	9
3.1	The continuous mixture of kernels	9
3.2	Scale decomposition	11
3.3	Restriction to a Finite Number of Scales	14
4	Conclusion and outlook	15
	References	16

Abstract

We develop a multi-scale theory for group of diffeomorphisms based on previous works [BGBHR11, RVW⁺11, SNLP11, SLNP11]. The purpose of the paper is to develop in details a variational approach for multi-scale on diffeomorphisms and to generalize to several scales the semi-direct product of group representation. We also show that the approaches presented in [SNLP11, SLNP11] and the mixture of kernels of [RVW+11] are equivalent.

1 Introduction

In this paper, we develop a multi-scale theory for groups of diffeomorphisms in the context of image registration. The setting of *large deformation diffeomorphic matching* has been introduced in seminal papers [Tro98, DGM98] and this approach has been applied in the field of computational anatomy [GM98]. The initial problem deals with the diffeomorphic registration of two given biomedical images or shapes. An important aspect of this model is the use of Reproducing Kernel Hilbert Spaces (RKHS) of vector valued functions to define the Lie algebra of the group of diffeomorphisms. Although our context is very different

from kernel based methods for machine learning [HTF09], some analogies may be worth noticing. We now present the model developed in [BMTY05].

Definition 1.1. Let Ω be a domain in \mathbb{R}^d . An admissible Reproducing Kernel Hilbert Space H of vector fields is a Hilbert space of $C^1(\Omega)$ vector fields such that there exists a positive constant M , s.t. for every $v \in H$ the following inequality holds

$$\|v\|_{1,\infty} \leq M_s \|v\|_H \quad (1)$$

Remark 1.2. The reproducing kernel associated with the space H is given by $K(x, y) = (\delta_x, K\delta_y)$, where δ_x, δ_y are the pointwise evaluation maps which are linear forms by assumption on the space H . The kernel completely defines the space H .

The diffeomorphism group G associated with the RKHS H is given by $\{\varphi_1 \mid v \in L^2([0, 1], H)\}$, where φ_t is the flow of v , *i.e.*

$$\begin{cases} \partial_t \varphi_t = v(t) \circ \varphi_t \\ \varphi_0 = \text{Id} . \end{cases} \quad (2)$$

The diffeomorphic matching problem is the minimization of the functional

$$\mathcal{F}(v) = \int_0^1 \|v(t)\|_H^2 dt + d(\varphi_1 \cdot q_0, q_{\text{target}}), \quad (3)$$

where $v \in L^2([0, 1], H)$ and q_0, q_{target} are objects of interest such as groups of points, measures, currents or images. The action of the group G on the objects space is denoted by $\varphi \cdot q$, where φ is an element of the group and q is an object. The distance function d that enforces the matching accuracy is usually taken to be the square of the norm if the objects live in a normed vector space, *e.g.* for images one would use the L^2 norm, $d(q_0, q_1) = \int_{\Omega} |q_0(x) - q_1(x)|^2 dx$. This minimization problem enables us to match images via geodesics on the group G , if we endow G with the right-invariant metric obtained by translating the inner product $\langle \cdot, \cdot \rangle_H$ on the Lie algebra H to the other tangent spaces. More importantly by its action on the space of images, the right-invariant metric on the group induces a metric on the orbits of the image space and the final deformation is completely encoded in the so-called initial momentum [VMTY04]. This initial momentum has the same dimension as the image. Since it is an element of a linear space, statistics can also be done on it [SFP⁺10].

A RKHS corresponding to a Gaussian kernel is commonly used in practice. The choice of its standard deviation σ is an important problem and describes a trade-off between the smoothness of the deformation and the matching accuracy. Indeed, a large standard deviation produces very smooth deformations with a poor matching accuracy of the structures having a size smaller than σ . On the contrary, a small standard deviation results in a good matching accuracy but the deformations may present undesirable very large Jacobians [RVW⁺11]. It is therefore a natural step to introduce a mixture of Gaussian kernels with different standard deviations. In [RVW⁺11], the authors show that such kernels outperform single Gaussian kernels, when registering images containing features of interest at several scales simultaneously: they provide a good matching quality, while keeping the diffeomorphisms smooth. Naturally, this introduces

more parameters in the algorithm, which need to be tuned. Practical insight about how to parametrize the scales and weights of multiple kernels are given in [RVW⁺11]. The idea of using a mixture of kernels for matching is directly connected to [BGBHR11], where it is proven that there is an equivalence between the matching with a sum of two kernels and the matching via a semi-direct product of two groups.

The work on the metric underlying the LDDMM methodology [BMTY05] has also been followed up by [SNLP11, SLNP11], where the authors introduce the notion of a bundle of kernels and argue that this general framework can be used to deal with multi-scale LDDMM. By passing, we prove that their approach reduces to the mixture of kernels. We give a self-contained and simple proof of this result based on Lagrange multipliers and we develop an extension of [BGBHR11], first to a finite number of scales and then to a continuum of scales. The paper is divided into three parts: the first part focuses on a finite number of scales while the second treats the case of a continuum of scales. The last part of the paper is devoted to numerical simulations, where we show in particular the decomposition on the given scales of the optimized diffeomorphism.

2 A finite number of scales

2.1 The finite mixture of kernels

For the sake of simplicity, we first treat the case of a finite set of admissible Hilbert spaces H_i for $i = 1, \dots, k$. Denoting $H = H_1 + \dots + H_k$, the norm proposed in [SNLP11] as well as in [BGBHR11] is defined by

$$\|v\|_H^2 = \inf \left\{ \sum_{i=1}^k \|v_i\|_{H_i}^2 \mid \sum_{i=1}^k v_i = v \right\}. \quad (4)$$

The following lemma is the main argument to prove the equivalence between the approaches of [SNLP11] and [RVW⁺11]. This lemma is an old result that can be found in [Aro50]. However, we present for the sake of completeness a simple proof based on the Lagrange multiplier rule. Moreover, if one wants to skip the technical details of the proof, the formal application of the Lagrange multiplier rule gives the result immediately. We outline the formal proof: If one has to minimize the norm defined in Formula (4) then one can introduce a Lagrange multiplier p and obtain a stationary point of the Lagrangian

$$L_v(v_1, \dots, v_n, p) = \sum_{i=1}^k \|v_i\|_{H_i}^2 + (p, v - \sum_{i=1}^k v_i) \quad (5)$$

which gives $v_i = K_i p$ and

$$v = \sum_{i=1}^k K_i p. \quad (6)$$

Hence, it gives a heuristic argument why the problem of optimizing at several scale simultaneously reduces to a mixture of the kernel.

Lemma 2.1. *The formula (4) defines a scalar product on H which makes H a RKHS and its associated kernel is $K := \sum_{i=1}^n K_i$, where K_i denotes the kernel of the space H_i .*

Proof. Let $x \in \Omega$, $\alpha \in \mathbb{R}^d$ and δ_x^α be the pointwise evaluation defined by $\delta_x^\alpha(v) \doteq \langle v(x), \alpha \rangle_{\mathbb{R}^d}$. By hypothesis on each H_i , δ_x^α is a linear form which implies that $Ev_{(x,\alpha)} : (v_i) \in \bigoplus_{i=1}^k H_i \mapsto \sum_{i=1}^k \delta_x^\alpha(v_i) \in \mathbb{R}$ is also a linear form on $\bigoplus_{i=1}^k H_i$. Now, for any v that can be written as $v = \sum_{i=1}^k u_i$ for $(u_i) \in \bigoplus_{i=1}^k H_i$, there exists a unique $(v_i) \in \bigoplus_{i=1}^k H_i$ minimizing the functional $N((v_i)_{i=1\dots k}) = \frac{1}{2} \sum_{i=1}^k \|v_i\|_{H_i}^2$ and satisfying $\sum_{i=1}^k v_i = v$. This unique element is given by the projection theorem for Hilbert space [Bre83] $V := \left(\bigcap_{(x,\alpha) \in \Omega \times \mathbb{R}^d} Ev_{(x,\alpha)}^{-1}(\{0\}) \right)^\perp$ which is a closed non-empty subspace in $\bigoplus_{i=1}^k H_i$. Therefore H is isometric to V and hence it is a Hilbert space.

In order to identify the kernel of H , we apply the Lagrange multiplier rule. An optimal $(v_i) \in \bigoplus_{i=1}^k H_i$ corresponds to a stationary point of the augmented functional on H^*

$$\tilde{N}(p, (v_i)_{i=1\dots k}) = \frac{1}{2} \sum_{i=1}^k \|v_i\|_{H_i}^2 + \left(p, v - \sum_{i=1}^k v_i \right)_{H^*, H}. \quad (7)$$

Remark that the norm $\|\cdot\|_H$ makes the injection $j_i : H_i \hookrightarrow H$ continuous and as a consequence $j_i^* : H^* \rightarrow H_i^*$ is defined by duality. Therefore the pairing (p, v_i) is well defined in Formula (7). Then, at a stationary point $(p, (v_i)_{i=1\dots k})$ we have

$$\begin{cases} v_i = K_i(p) \text{ for } i = 1 \dots k \\ v = \sum_{i=1}^k v_i = \sum_{i=1}^k K_i(p). \end{cases} \quad (8)$$

Note that in the previous formula, we could have written the heavier notation $v_i = K_i(j_i^* p)$ to be more precise. Taking the dual pairing of the last equation with p we get

$$(p, v) = \sum_{i=1}^k \|v_i\|_{H_i}^2 = \sum_{i=1}^k (p, K_i p) = \|v\|_H^2,$$

which implies that the Riesz isomorphism between H^* and H is given by the map $p \in H^* \mapsto \sum_{i=1}^k K_i(p) \in H$. Moreover, we have $\bigcap_{i=1}^k H_i^* \subset H^*$: For $p \in \bigcap_{i=1}^k H_i^*$ we have,

$$|(p, v)| \leq \sum_{i=1}^k |(p, v_i)| \leq \sum_{i=1}^k \|p\|_{H_i^*} \|v_i\|_{H_i} \stackrel{C.S.}{\leq} \sqrt{\sum_{i=1}^k \|p\|_{H_i^*}^2} \sqrt{\sum_{i=1}^k \|v_i\|_{H_i}^2},$$

which is true for any decomposition of v so that

$$|(p, v)| \leq \sqrt{\sum_{i=1}^k \|p\|_{H_i^*}^2} \|v\|_H. \quad (9)$$

Since $\delta_x^\alpha \in \bigcap_{i=1}^k H_i^* \subset H^*$, H is a RKHS and its kernel is $K = \sum_{i=1}^k K_i$. \square

We now define the isometric injection of H in $\bigoplus_{i=1}^k H_i$.

Definition 2.2. We denote by $\pi : H \mapsto \bigoplus_{i=1}^k H_i$ the map defined by $\pi(v) = (K_i(K^{-1}(v)))_{i=1\dots k}$.

The non-linear version of this multi-scale approach to diffeomorphic matching problems is the minimization of

$$\mathcal{E}(v) = \int_0^1 \sum_{i=1}^k \|v_i(t)\|_{H_i}^2 dt + d(\varphi_1 \cdot q_0, q_{target}), \quad (10)$$

defined on $\bigoplus_{i=1}^k H_i$. Recall that φ_t is the flow generated by $v(t) := \sum_{i=1}^k v_i(t)$. The direct consequence of Lemma 2.1 is the following proposition.

Proposition 2.3. The minimization of \mathcal{E} reduces to the minimization of

$$\mathcal{F}(v) = \int_0^1 \|v(t)\|_H^2 dt + d(\varphi_1 \cdot q_0, q_{target}). \quad (11)$$

Proof. Obviously, the minimization of \mathcal{F} is the minimization of \mathcal{E} restricted to $\pi(H)$. Remark first that for any $(v_i) \in L^2([0, 1], \bigoplus_{i=1}^k H_i)$ then $\pi((v_i)) \in L^2([0, 1], \bigoplus_{i=1}^k H_i)$. Denoting $v = \sum_{i=1}^k v_i$ we have $\|\pi(v)\|_{L^2} \leq \|(v_i)\|_{L^2}$ with equality if and only if $\pi(v) = (v_i)$. Therefore, if $(v_i) \in L^2([0, 1], \bigoplus_{i=1}^k H_i)$ is a minimizer of \mathcal{E}

$$\mathcal{E}(\pi(v)) \leq \mathcal{E}((v_i)) \quad (12)$$

which implies $\pi(v) = (v_i)$ and the result. \square

Remark 2.4. To a minimizing path $v(t) \in L^2([0, 1], H)$ corresponds a minimizing path in $\bigoplus_{i=1}^k H_i$ via the map π . In other words, the optimal path can be decomposed on the different scales using each kernel K_i present in the reproducing kernel of H , $K = \sum_{i=1}^k K_i$.

2.2 Iterated semi-direct product

Until now, the scales have been introduced only on the Lie algebra of the diffeomorphism group and an important question is how to decompose the flow of diffeomorphisms according to these scales. An answer in the case of two scales is given in [BGBHR11], where the flow is decomposed with the help of a semi-direct product of groups and the whole transformation is encoded in a large-scale deformation and a small-scale one. The underlying idea is to represent the diffeomorphism flow $\varphi(t)$ by a pair $(\psi_1(t), \psi_2(t))$ where $\psi_1(t)$ is given by the flow of the vector field $v_1(t)$ and $\psi_2(t) := \varphi(t) \circ (\psi_1(t))^{-1}$. Remark in particular that $\psi_2(t)$ is not the flow of $v_2(t)$. More precisely, we have

$$\begin{cases} \partial_t \psi_1(t) = v_1(t) \circ \psi_1(t) \\ \partial_t \psi_2(t) = (v_1(t) + v_2(t)) \circ \psi_2(t) - \text{Ad}_{\psi_2(t)} v_2(t) \circ \psi_2(t). \end{cases} \quad (13)$$

Interestingly, as shown in [BGBHR11], this decomposition of the diffeomorphism flow corresponds to a semi-direct product of group. This framework can be generalized to a finite number of scales as follows.

Given n scales, we want to represent $\varphi(t)$ by an n -tuple $(\psi_1(t), \dots, \psi_n(t))$, with $\psi_1(t)$ corresponding to the finest scale and $\psi_n(t)$ to the coarsest scale. The geometric construction underlying the decomposition into multiple scales is the semidirect product, introduced in the following lemma.

Lemma 2.5. *Let $G_1 \supseteq G_2 \supseteq \dots \supseteq G_n$ be a chain of Lie groups. The n -fold semidirect product is given by the multiplication*

$$(g_1, \dots, g_n) \cdot (h_1, \dots, h_n) = (g_1 c_{g_2 \dots g_n} h_1, g_2 c_{g_3 \dots g_n} h_2, \dots, g_n h_n)$$

with $c_g h = ghg^{-1}$ denoting conjugation. Then given a right-invariant velocity field $v(t) = (v_1(t), \dots, v_n(t))$ the curve $g(t)$ is reconstructed via

$$\partial_t g_k(t) = \left(v_k(t) + (\text{Id} - \text{Ad}_{g_k(t)}) \sum_{i=1}^{k-1} v_i(t) \right) g_k(t) , \quad (14)$$

if $k \geq 2$ and $\partial_t g_1(t) = v_1(t)g_1(t)$ which is the semidirect product equivalent of $\partial_t g(t) = v(t)g(t)$. We shall denote this semidirect product by

$$G_1 \rtimes \dots \rtimes G_n$$

to emphasize that each subproduct $G_k \rtimes \dots \rtimes G_n$ is a normal subgroup of the whole product.

Proof. Verifying the axioms for the group multiplication is a straight forward, if slightly longer calculation. The inverse is given by

$$(g_1, \dots, g_n)^{-1} = (c_{(g_2 \dots g_n)^{-1}} g_1^{-1}, \dots, g_n^{-1}) .$$

The right hand side of equation (14) can be obtained by differentiation the group multiplication at the identity, *i.e.* computing $\partial_t(h(t) \cdot g)|_{t=0}$ with g fixed, $h(0) = e$ and $\partial_t h(t)|_{t=0} = v$. \square

In our case the group G_k is the diffeomorphism group $\text{Diff}_{K_k}(\Omega)$ generated by vector fields in the space H_{K_k} corresponding to the kernel K_k . The subgroup condition $\text{Diff}_{K_k}(\Omega) \supseteq \text{Diff}_{K_{k+1}}(\Omega)$ is satisfied, if we impose the corresponding condition $H_{K_k} \supseteq H_{K_{k+1}}$ on the spaces of vector fields, which we will assume from now on.

Starting from an n -tuple $v_1(t), \dots, v_n(t)$ of vector fields, we can reconstruct the diffeomorphisms at each scale via

$$\partial_t \psi_k(t) = \left(v_k(t) + (\text{Id} - \text{Ad}_{\psi_k(t)}) \sum_{i=1}^{k-1} v_i(t) \right) \circ \psi_k(t) \quad (15)$$

as in Lemma 2.5. We can also sum over all scales to form $v(t) = \sum_{k=1}^n v_k(t)$ and compute the flow $\varphi(t)$ of $v(t)$. Then a simple calculation shows that

$$\varphi(t) = \psi_1(t) \circ \dots \circ \psi_n(t) .$$

This construction can be summarized by the following commuting diagram

$$\begin{array}{ccc} (v_1(t), \dots, v_n(t)) & \xrightarrow{v(t)=\sum v_k(t)} & v(t) \\ \text{via (15)} \downarrow & & \downarrow \partial_t \varphi(t)=v(t) \circ \varphi(t) \\ (\psi_1(t), \dots, \psi_n(t)) & \xrightarrow{\varphi(t)=\psi_1(t) \circ \dots \circ \psi_n(t)} & \varphi(t) \end{array} \quad (16)$$

We can now formulate several equivalent versions of LDDMM matching with multiple scales. The most straight-forward way is to do matching with a kernel, which is a sum of kernels of different scales. This is the approach considered in [RVW⁺11].

Definition 2.6 (LDDMM with Sum-of-Kernels). Registering I_0 to I_{targ} is done by finding the minimizer $v(t)$ of

$$\frac{1}{2} \int_0^1 \|v(t)\|_K^2 dt + \frac{1}{2\sigma^2} \|\varphi(1).I_0 - I_{\text{targ}}\|^2 ,$$

where $K = \sum_{i=0}^n K_i$ is a sum of n kernels.

The corresponding simultaneous multiscale registration problem, where one assigns to each scale a separate vector field is a special case of the kernel bundle method proposed in [SNLP11] and [SLNP11].

Definition 2.7 (Simultaneous Multiscale Registration). Registering I_0 to I_{targ} is done by finding the minimizing n -tuple $(v_1(t), \dots, v_n(t))$ of

$$\frac{1}{2} \sum_{i=1}^n \int_0^1 \|v_i(t)\|_{K_i}^2 dt + \frac{1}{2\sigma^2} \|\varphi(1).I_0 - I_{\text{targ}}\|^2 ,$$

where $\varphi(t)$ is the flow of the vector field $v(t) = \sum_{i=1}^n v_i(t)$.

The geometric version of the multiscale registration not only uses separate vector fields for each scale, but also decomposes the diffeomorphisms according to scales and can be defined as follows.

Definition 2.8 (LDDMM with a Semidirect Product). Registering I_0 to I_{targ} is done by finding the minimizing n -tuple $(v_1(t), \dots, v_n(t))$ of

$$\frac{1}{2} \sum_{i=1}^n \int_0^1 \|v_i(t)\|_{K_i}^2 dt + \frac{1}{2\sigma^2} \|\varphi(1).I_0 - I_{\text{targ}}\|^2 ,$$

where $\varphi(t) = \psi_1(t) \circ \dots \circ \psi_n(t)$ and $\psi_i(t)$ is defined via (15).

Problem 2.8 can be obtained from the abstract framework in [BGBHR11] by considering the action

$$\begin{aligned} (\text{Diff}_{K_1}(\Omega) \rtimes \dots \rtimes \text{Diff}_{K_n}(\Omega)) \times V &\rightarrow V \\ ((\psi_1, \dots, \psi_n), I) &\mapsto I \circ \psi_n^{-1} \circ \dots \circ \psi_1^{-1} \end{aligned} \quad (17)$$

of the semidirect product on the space of images.

Theorem 2.9. *The matching problems 2.6, 2.7 and 2.8 are all equivalent.*

Proof. The equivalence of problems 2.6 and 2.7 follows from Proposition 2.3 while the equivalence of problems 2.7 and 2.8 follows from the diagram (16). For the case $n = 2$ the proof can be found in more details in [BGBHR11]. \square

This construction can also be generalized to a continuum of scales as introduced in [SNLP11] and [SLNP11]. We present in the next section a more general framework to deal with such continuous multi-scale decompositions.

2.3 The Order Reversed

The action (17) of the semidirect product from Lemma 2.5 proceeds by deforming the image with the coarsest scale diffeomorphism first and with the finest scale diffeomorphism last. However, it is also possible to reverse this order and to act with the finest scale diffeomorphisms first. We will see that this approach also corresponds to a semidirect product and is equivalent to the other ordering of scales via a change of coordinate. The reason to expand on this here is that it is better suited to be generalized to a continuum of scales.

In this section we will assume that the group G_1 contains the deformations of the coarsest scale and G_n those of the finest scale. The corresponding semidirect product is described in the following lemma.

Lemma 2.10. *Let $G_1 \subseteq G_2 \subseteq \dots \subseteq G_n$ be a chain of Lie groups. The n -fold semidirect product is given by the multiplication*

$$(g_1, \dots, g_n) \cdot (h_1, \dots, h_n) = (g_1 h_1, c_{h_1^{-1}} g_2 h_2, \dots, c_{(h_1 \dots h_{n-1})^{-1}} g_n h_n)$$

with $c_g h = ghg^{-1}$ denoting conjugation. Then given a right-invariant velocity field $v(t) = (v_1(t), \dots, v_n(t))$ the curve $g(t)$ is reconstructed via

$$\partial_t g_k(t) = \text{Ad}_{(g_1(t) \dots g_{k-1}(t))^{-1}} v_k(t),$$

which is the semidirect product equivalent of $\partial_t g(t) = v(t)g(t)$. We shall denote this product by

$$G_1 \ltimes \dots \ltimes G_n$$

to emphasize that each subproduct $G_1 \ltimes \dots \ltimes G_k$ from the left is a normal subgroup of the whole product.

Proof. This lemma can be proven in the same way as Lemma 2.5. \square

These semidirect products defined in Lemmas 2.5 and 2.10 are equivalent as shown by the following lemma

Lemma 2.11. *The map*

$$\varphi : \begin{cases} G_1 \ltimes \dots \ltimes G_n & \rightarrow & G_n \ltimes \dots \ltimes G_1 \\ (g_1, \dots, g_n) & \mapsto & (g_n, \dots, c_{(g_{n+2-k} \dots g_n)^{-1}} g_{n+1-k}, \dots, c_{(g_2 \dots g_n)^{-1}} g_1) \end{cases}$$

is a group homomorphism between the two semidirect products and its derivative at the identity is given by

$$T_e \varphi : \begin{cases} \mathfrak{g}_1 \ltimes \dots \ltimes \mathfrak{g}_n & \rightarrow & \mathfrak{g}_n \ltimes \dots \ltimes \mathfrak{g}_1 \\ (v_1, \dots, v_n) & \mapsto & (v_n, \dots, v_{n+1-k}, \dots, v_1) \end{cases}$$

Proof. Direct computation. \square

The map φ can be seen as one side of the following commuting triangle

$$\begin{array}{ccc} G_1 \ltimes \dots \ltimes G_n & \xrightarrow{\varphi} & G_n \ltimes \dots \ltimes G_1 \\ & \searrow T_1 \quad \swarrow T_2 & \\ & G_1 \times \dots \times G_n & \end{array}$$

The maps

$$\begin{aligned} T_1(g_1, \dots, g_n) &= (g_1 \cdots g_n, \dots, g_n) \\ T_2(g_1, \dots, g_n) &= (g_n, g_{n-1}g_n, \dots, g_1 \cdots g_n) \end{aligned}$$

are group homomorphisms from the corresponding semidirect products into the direct product $G_1 \times \dots G_n$. They can be regarded as trivializations of the semidirect product in the special case that the factors form a chain of subgroups.

We will now assume that we are given n kernels K_1, \dots, K_n such that the inclusions $H_{K_i} \subseteq H_{K_{i+1}}$, *i.e.* K_1 represents the coarsest scale and K_n the finest one. Note that the inclusions are reversed as compared to Section 2.2. The registration problem is now defined as follows.

Definition 2.12 (LDDMM with the alternative Semidirect Product). Registering I_0 to I_{targ} is done by finding the minimizing n -tuple $(v_1(t), \dots, v_n(t))$ of

$$\frac{1}{2} \sum_{i=1}^n \int_0^1 \|v_i(t)\|_{K_i}^2 dt + \frac{1}{2\sigma^2} \|\varphi(1) \cdot I_0 - I_{\text{targ}}\|^2 ,$$

where $\varphi(t) = \psi_1(t) \circ \dots \circ \psi_n(t)$ and $\psi_k(t)$ is defined via

$$\partial_t \psi_k(t) = (\text{Ad}_{(\psi_1(t) \circ \dots \circ \psi_{k-1}(t))^{-1}} v_k(t)) \circ \psi_k(t) .$$

To see that problems 2.8 and 2.12 are equivalent note that the following diagram commutes

$$\begin{array}{ccc} G_1 \rtimes \dots \rtimes G_n & \xrightarrow{\varphi} & G_n \rtimes \dots \rtimes G_1 \\ & \searrow \varphi = \psi_1 \dots \psi_n & \swarrow \varphi = \psi_n \dots \psi_1 \\ & G_1 & \end{array}$$

and that $T_e \varphi$ merely reverses the order of the vector fields $(v_1(t), \dots, v_n(t))$. We will see in Section 3.2 that this version of the semidirect product is better suited to be generalized to a continuum of scales.

3 Extension to a continuum of scales

3.1 The continuous mixture of kernels

In this section, we define the multi-scale approach for a continuum of scales. First, we introduce the necessary analytical framework and useful results.

Definition 3.1. Let Ω be a domain in \mathbb{R}^d . An *admissible bundle* (\mathcal{H}, λ) is a one-parameter family of Reproducing Kernel Hilbert Spaces of $C^1(\Omega)$ vector fields H_s indexed by $s \in \mathbb{R}_+^*$ such that for any s , there exists a positive constant M_s s.t. for every $v \in H_s$

$$\|v\|_{1,\infty} \leq M_s \|v\|_{H_s} . \quad (18)$$

The map $\mathcal{K} : \mathbb{R}_+^* \times \mathbb{R}^d \times \mathbb{R}^d \ni (s, x, y) \mapsto K_s(x, y) \in L(\mathbb{R}^d)$ is assumed to be Borel measurable. In addition to this set of linear spaces, λ is a Borel measure

on \mathbb{R}_+^* such that

$$\int_{\mathbb{R}_+^*} M_s^2 \, d\lambda(s) < +\infty, \quad (19)$$

where $s \rightarrow M_s$ is also assumed to be Borel measurable.

Remark 3.2. • The hypothesis of the definition may not be optimal to obtain the needed property, but this context is already large enough for applications.

• Remark that no inclusion is a priori required between the linear spaces H_s . However the typical example is given by the usual scale-space *i.e.* \mathcal{H}_s defined by its Gaussian kernel $e^{-\frac{\|x-y\|^2}{2s^2}}$. In this case, there exists an obvious tower $H_s \subset H_t$ for $s > t > 0$.

• We have $\int_{\mathbb{R}_+^*} K_s(x, y) \, d\lambda(s) \leq \int_{\mathbb{R}_+^*} M_s^2 \, d\lambda(s)$. This comes from Cauchy-Schwarz's inequality to get $\langle \alpha, K_s(x, y) \beta \rangle \leq \|\delta_x^\alpha\|_{H_s} \|\delta_y^\beta\|_{H_s}$ and the fact that $\|\delta_x^\alpha\|_{H_s} \leq |\alpha| M_s$.

Mimicking Section 2, we consider the set of vector-valued functions defined on $\mathbb{R}_+^* \times \Omega$, namely denoting μ the Lebesgue measure on Ω ,

$$V := \left\{ f \in L^2(\mathbb{R}_+^* \times \Omega, \lambda \otimes \mu) \left| \begin{array}{l} (1) \forall x \in \Omega \quad s \rightarrow f(s, x) \text{ is measurable.} \\ (2) s \rightarrow \|f(s, \cdot)\|_{H_s} \text{ is measurable.} \\ (3) \|f\|_V^2 = \int_{\mathbb{R}_+^*} \|f(s, \cdot)\|_{H_s}^2 \, d\lambda(s) < +\infty. \end{array} \right. \right\} \quad (20)$$

It is rather standard to prove that V is a Hilbert space for the norm introduced in (20). Remark that V contains all the functions $(s, x) \rightarrow K_s(x, \cdot)$ for any $x \in \Omega$. Directly from the assumptions on the space V , we can define the set of functions

$$H := \left\{ x \rightarrow \int_{\mathbb{R}_+^*} v_s(x) \, d\lambda(s) \left| v \in V \right. \right\}. \quad (21)$$

Then, the generalization of Lemma 2.1 [Sai88, Sch64] reads in our situation:

Theorem 3.3. *The space H can be endowed with the following norm: For any $g \in H$,*

$$\|g\|_H^2 = \inf_{f \in V} \int_{\mathbb{R}_+^*} \|f_s\|_{H_s}^2 \, d\lambda(s), \quad (22)$$

for f satisfying the constraint $g = \int_{\mathbb{R}_+^*} f_s \, d\lambda(s)$. This norm turns H into a RKHS whose kernel is $\mathcal{K}(x, y) = \int_{\mathbb{R}_+^*} K_s(x, y) \, d\lambda(s)$.

In our case, the hypothesis on the bundle (\mathcal{H}, λ) implies that H is an admissible RKHS.

Proof. As mentioned above, the map

$$\begin{aligned} \pi : V &\mapsto H \\ v &\mapsto \int_{\mathbb{R}_+^*} v_s \, d\lambda(s) \end{aligned}$$

is well defined and so is $\pi^* : H^* \mapsto V^*$. Using Cauchy-Schwarz's inequality, we have

$$\left| \int_{\mathbb{R}_+^*} v_s(x) \, d\lambda(s) \right| \leq \|f_s\|_{H_s} \sqrt{\int_{\mathbb{R}_+^*} M_s^2 \, d\lambda(s)}$$

$$|\pi(v)(x)| \leq |\pi(v)| \sqrt{\int_{\mathbb{R}_+^*} M_s^2 \, d\lambda(s)}. \quad (23)$$

Hence, the evaluation map $\delta_x \circ \pi : v \in V \rightarrow \int_{\mathbb{R}_+^*} v_s(x) \, d\lambda(s)$ is continuous on V . Therefore the map π is continuous for the product topology on $(\mathbb{R}^d)^\Omega$ and its kernel is a closed subspace denoted by \tilde{H} . Applying the projection theorem on \tilde{H}^\perp , for any $u \in H$ there exists a unique \tilde{u} such that $\|\tilde{u}\|_V^2 = \inf_{v \in V} \{\|v\|_V^2 \mid \pi(v) = u\}$. Hence, Equation (22) defines a norm on H for which $\pi|_{\tilde{H}^\perp}$ is an isometry. In particular H is a Hilbert space.

Remark that Inequality (23) means that H is a RKHS. We can now apply the Lagrange multiplier rule in a Hilbert space on the following functional

$$\frac{1}{2} \|v\|_V^2 + (p, u - \pi(v))_{H^*, H}. \quad (24)$$

where $(u, p) \in H \times H^*$ and $v \in V$. For a stationary point, we obtain $\pi(v) = u$, $K_s \pi^* p(s) = v_s$, λ a.e. The Riesz isomorphism is then given by

$$\mathbf{K} : p \in H^* \rightarrow \int_{\mathbb{R}_+^*} K_s \pi^*(p)(s) \, d\lambda(s) \in V, \quad (25)$$

and the kernel function is given by

$$(\delta_x, \mathbf{K} \delta_y) = \int_{\mathbb{R}_+^*} \mathcal{K}_s(x, y) \, d\lambda(s). \quad (26)$$

□

Remark 3.4. Importantly, the hypothesis on the bundle (\mathcal{H}, λ) implies that H is an admissible RKHS since we can apply a theorem of differentiation under the integral sign. By application of Cauchy-Schwarz's inequality to the integral defining $\pi(v)$ for any $v \in V$, we have that

$$\|\pi(v)\|_{1,\infty}^2 \leq \int_{\mathbb{R}_+^*} M_s^2 \, d\lambda(s) \int_{\mathbb{R}_+^*} \|v_s\|_{H_s^*}^2 \, d\lambda(s). \quad (27)$$

3.2 Scale decomposition

In this section we will generalize the ideas of Section 2.2 from a finite sum of kernels to a continuum of scales. We will make some more assumptions to the general setting introduced in Section 3.1.

Assumption. We assume that the measure $\lambda(s)$ is the Lebesgue measure on the finite interval $[0, 1]$ and that the family H_s of RKHS is ordered by inclusion

$$H_s \subseteq H_t \text{ for } s \leq t$$

This assumption might be relaxed a little bit. The main restriction is that we don't want the measure λ to contain any singular parts. As long as $\lambda(s)$ is absolutely continuous with respect to the Lebesgue measure, *i.e.* it can be represented via a density $\lambda(s) = f(s) dx$, the same construction can be carried out. The ordering of the inclusions corresponds to that in Section 2.3.

As in the discrete setting we can formulate the two image matching problems. The first is a direct generalization of problem 2.6 to a continuum of scales.

Definition 3.5 (LDDMM with an Integral Kernel). Registering I_0 to I_{targ} is done by finding the minimizer $v(t)$ of

$$\frac{1}{2} \int_0^1 \|v(t)\|_K^2 dt + \frac{1}{2\sigma^2} \|\varphi(1).I_0 - I_{\text{targ}}\|^2 ,$$

where $K = \int_0^1 K_s ds$ is the integral over the scales K_s .

The other problem associates to each scale a separate vector field. It was proposed in [SNLP11, SLNP11], where it was called registration with a kernel bundle. The term kernel bundle refers to the one-parameter family $\mathcal{H} = (H_s)_{s \in [0,1]}$ of RKHS.

Definition 3.6 (LDDMM with a Kernel Bundle). Registering I_0 to I_{targ} is done by finding the one-parameter family $v_s(t)$ of vector fields, which minimizes

$$\frac{1}{2} \int_0^1 \int_0^1 \|v_s(t)\|_{K_s}^2 ds dt + \frac{1}{2\sigma^2} \|\varphi(1).I_0 - I_{\text{targ}}\|^2 ,$$

where $\varphi(t)$ is the flow of the vector field $v(t) = \int_0^1 v_s(t) ds$.

These two problems are equivalent, as will be shown in Theorem 3.11. As a next step we want to obtain a geometric reformulation of the registration problem similar to 2.8 or 2.12. The goal of this reformulation is to decompose the minimizing flow of diffeomorphisms $\varphi(t)$, such that the effect of each scale becomes visible. In order to do this decomposition we define

$$\psi_s(t) \text{ to be the flow of } \int_0^s v_r(t) dr . \quad (28)$$

The following theorem allows us to interchange time and scale in the flow $\psi_s(t)$.

Theorem 3.7. *For each fixed t , the one-parameter family $s \mapsto \psi_s(t)$ is the flow of*

$$\text{Ad}_{\psi_s(t)} \int_0^t \text{Ad}_{\psi_s(r)^{-1}} v_s(r) dr .$$

To prove this theorem we will use the following lemma.

Lemma 3.8. *Let $u(s, t, x)$ and $v(s, t, x)$ be two-parameter families of vector fields which are C^2 in the (s, t) -variables and C^1 in x . If they satisfy*

$$\partial_s u(s, t, x) - \partial_t v(s, t, x) = [v(s, t), u(s, t)](x)$$

and if $v(s, 0) \equiv 0$ for all s , then the flow of $u(s, \cdot)$ for fixed s coincides with the flow of $v(\cdot, t)$ for fixed t .

Proof. Denote by $a_s(t)$ the flow of $u(s, \cdot)$ in t . Then

$$\begin{aligned}\partial_t \partial_s a_s(t) &= \partial_s \partial_t a_s(t) = \partial_s (u(s, t) \circ a_s(t)) \\ &= \partial_s u(s, t) \circ a_s(t) + Du(s, t, a_s(t)) \cdot \partial_s a_s(t) \\ &= \partial_t v(s, t) \circ a_s(t) + [v(s, t), u(s, t)] \circ a_s(t) + Du(s, t, a_s(t)) \cdot \partial_s a_s(t) \\ &= \partial_t (v(s, t) \circ a_s(t)) + Du(s, t, a_s(t)) \cdot (\partial_s a_s(t) - v(s, t) \circ a_s(t))\end{aligned}$$

This implies that $b_s(t) := \partial_s a_s(t) - v(s, t) \circ a_s(t)$ is the solution of the ODE

$$\partial_t b_s(t) = Du(s, t, a_s(t)) \cdot b_s(t) . \quad (29)$$

Since for $t = 0$ we have $b_s(0) = \partial_s a_s(0) - v(s, 0) \circ a_s(t) = 0$, it follows that $b_s(t) \equiv 0$ is the unique solution of (29). This means that

$$\partial_s a_s(t) = v(s, t) \circ a_s(t) ,$$

i.e. the flows of $u(s, \cdot)$ in t and of $v(\cdot, t)$ in s coincide. \square

Proof of Theorem. We apply Lemma 3.8 to the vector fields

$$\int_0^s v_r(t) dr \text{ and } \text{Ad}_{\psi_s(t)} \int_0^t \text{Ad}_{\psi_s(r)^{-1}} v_s(r) dr .$$

We can differentiate Ad using the following rule

$$\partial_t \text{Ad}_{g(t)} u = [\partial_t g(t) g(t)^{-1}, \text{Ad}_{g(t)} u] .$$

Using this we can verify the compatibility condition

$$\begin{aligned}\partial_s \left(\int_0^s v_r(t) dr \right) - \partial_t \left(\text{Ad}_{\psi_s(t)} \int_0^t \text{Ad}_{\psi_s(r)^{-1}} v_s(r) dr \right) &= \\ &= v_s(t) - \left[\partial_t \psi_s(t) \psi_s(t)^{-1}, \text{Ad}_{\psi_s(t)} \int_0^t \text{Ad}_{\psi_s(r)^{-1}} v_s(r) dr \right] - v_s(t) \\ &= \left[\int_0^s v_s(t) dr, \text{Ad}_{\psi_s(t)} \int_0^t \text{Ad}_{\psi_s(r)^{-1}} v_s(r) dr \right] .\end{aligned}$$

The condition $v(s, 0) \equiv 0$ is trivially satisfied. This concludes the proof. \square

The main conclusion from Theorem 3.7 can be summarized in the corollary.

Corollary 3.9. *Let $\varphi(t)$ be the flow of the vector field $\int_0^1 v_s(t) ds$ and define $\eta(s)$ via the left-invariant velocity, i.e.*

$$\eta(s)^{-1} \partial_s \eta(s) = \int_0^1 \text{Ad}_{\varphi(t)^{-1}} v_s(t) dt .$$

Then we have

$$\varphi(1) = \eta(1) ,$$

i.e. $\eta(s)$ is the flow in scale of $\varphi(t)$.

Proof. Apply Theorem 3.7 with $s = 1$. \square

Corollary 3.9 gives us a way to decompose the matching diffeomorphism $\varphi(1)$ into separate scales. As we follow the flow $\eta(s)$, we add more and more scales, starting from the identity, when no scales are taken into account and finishing with $\varphi(1)$, which includes all scales. In this sense $\eta(t) \circ \eta(s)^{-1}$ contains the scale information for the scales in the interval $[s, t]$.

3.3 Restriction to a Finite Number of Scales

It is of interest to understand the relation between a continuum of scales and the case, where we have only a finite number. We will see, that it is possible to see the discrete case as an approximation of the continuous case.

Let us start with a family K_s of kernels with $s \in [0, 1]$, where the scales are ordered from the coarsest to the finest, *i.e.* $H_{K_s} \leq H_{K_t}$ for $s \leq t$ as before. Divide the interval $[0, 1]$ into n parts $0 = t_0 < \dots < t_n = 1$ and denote the intervals $I_k = [t_{k-1}, t_k]$. Define the space

$$L^2(\mathcal{H}) = \left\{ v \in \mathcal{H} : \int_0^1 \|v_s\|_{H_s}^2 ds < \infty \right\} ,$$

where \mathcal{H} was defined in Definition 3.1 to be a one-parameter family of vector fields $v_s \in H_s$. To each interval I_k corresponds a kernel $K_k = \int_{I_k} K_s ds$. The discrete sampling map

$$\Psi : \begin{cases} L^2(\mathcal{H}) & \rightarrow H_{K_1} \times \dots \times H_{K_n} \\ v_s & \mapsto (\int_{I_1} v_s ds, \dots, \int_{I_n} v_s ds) \end{cases}$$

discretizes v_s into n scales. Formally we can introduce a Lie bracket on the space $L^2(\mathcal{H})$ by defining

$$[u, v]_s = \left[u_s, \int_0^s v_r dr \right] + \left[\int_0^s u_r dr, v_s \right] . \quad (30)$$

Using this bracket the sampling map Ψ is a Lie algebra homomorphism as shown in the next theorem.

Theorem 3.10. *The sampling map Ψ is a Lie algebra homomorphism from the Lie algebra $L^2(\mathcal{H})$ with the bracket defined in (30) into the n -fold semidirect product with the bracket*

$$[(u_1, \dots, u_n), (v_1, \dots, v_n)] = ([u_1, v_1], \dots, [u_k, \sum_{i=1}^{k-1} v_i] + [\sum_{i=1}^{k-1} u_i, v_k] + [u_k, v_k], \dots)$$

Proof. Using the definitions we first compute

$$\begin{aligned} [\Psi(u), \Psi(v)]_k &= \left[\int_{t_{k-1}}^{t_k} u_s ds, \int_0^{t_{k-1}} v_s ds \right] + \left[\int_0^{t_{k-1}} u_s ds, \int_{t_{k-1}}^{t_k} v_s ds \right] + \\ &\quad + \left[\int_{t_{k-1}}^{t_k} u_s ds, \int_{t_{k-1}}^{t_k} v_s ds \right] \\ &= \left[\int_0^{t_k} u_s ds, \int_0^{t_k} v_s ds \right] - \left[\int_0^{t_{k-1}} u_s ds, \int_0^{t_{k-1}} v_s ds \right] , \end{aligned}$$

and then write the other side

$$\Psi([u, v])_k = \int_{t_{k-1}}^{t_k} \left[u_s, \int_0^s v_r dr \right] + \left[\int_0^s u_r dr, v_s \right] ds$$

Below we interchange the order of integration in the first summand and merely switch s and r in the second summand to obtain

$$\begin{aligned}
& \int_0^{t_k} \left[u_s, \int_0^s v_r \, dr \right] + \left[\int_0^s u_r \, dr, v_s \right] \, ds = \\
&= \int_0^{t_k} \int_0^s [u_s, v_r] \, dr \, ds + \int_0^{t_k} \int_0^r [u_s, v_r] \, ds \, dr \\
&= \int_0^{t_k} \int_r^{t_k} [u_s, v_r] \, ds \, dr + \int_0^{t_k} \int_0^r [u_s, v_r] \, ds \, dr \\
&= \int_0^{t_k} \int_0^{t_k} [u_s, v_r] \, ds \, dr \\
&= \left[\int_0^{t_k} u_s \, ds, \int_0^{t_k} v_s \, ds \right] .
\end{aligned}$$

Decomposing the integral into

$$\Psi([u, v])_k = \int_0^{t_k} \dots \, ds - \int_0^{t_{k-1}} \dots \, ds$$

finishes the proof. \square

Now we can show that all matching problems that we defined in the continuous case are equivalent.

Theorem 3.11. *The matching problems 3.5 and 3.6 are equivalent and using the sampling map Ψ they are also equivalent to the discrete problem 2.12.*

Proof. The first equivalence follows from Theorem 3.3. For the second equivalence note that problem 2.12 is equivalent to 2.6 and use Lemma 2.1. \square

The diffeomorphisms ψ_k at each scale, that were defined in 2.12 are also contained in the continuous setting. Note that $\psi_1(t) \circ \dots \circ \psi_k(t)$ was the flow of the vector field $v_1(t) + \dots + v_k(t)$ and using the map Ψ we have

$$v_1(t) + \dots + v_k(t) = \int_0^{t_k} v_s(t) \, ds .$$

Hence we obtain the identity $\psi_1(t) \circ \dots \circ \psi_k(t) = \psi_{t_k}(t)$, where ψ_{t_k} was defined in (28). In particular we retrieve

$$\psi_k(t) = \psi_{t_{k-1}}(t)^{-1} \circ \psi_{t_k}(t)$$

the scale decomposition of the discrete case as a continuous scale decomposition evaluated at specific points.

4 Conclusion and outlook

In this paper, we have extended the mixture of kernels presented in [RVW⁺11] to the continuous setting and we have given a variational interpretation of the matching problem. In particular, we have shown that the approaches presented in [SNLP11, SLNP11] and the mixture of kernels of [RVW⁺11] are equivalent.

Motivated by the mathematical development of the multi-scale approach to group of diffeomorphisms, we have extended the semi-direct product result of [BGBHR11] to a finite number of scales and also to a continuum of scales. Developing statistics in a multi-scale framework for the initial momentum is encouraged by previous results on the statistical use of each scales in [RVHR10]. To this aim, it would be very interesting question to obtain results about the sparsity of the scales used in the mixture of kernels in order to improve the statistical power on the initial momentum. In analogy with [KBS⁺09] but from the image registration point of view, [RVW⁺11] tends to favor a non-sparse description of the scales, which tends towards the continuum of scales. In this direction, a more theoretical approach to learning the parameters involved in the mixture of kernels still needs to be developed.

References

- [Aro50] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.
- [BGBHR11] M. Bruveris, F. Gay-Balmaz, D. Holm, and T. Ratiu. The momentum map representation of images. *Journal of Nonlinear Science*, 21:115–150, 2011. 10.1007/s00332-010-9079-5.
- [BMTY05] M. F. Beg, M. I. Miller, A. Trounev, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision*, 61(2):139–157, 2005.
- [Bre83] H. Brezis. *Analyse Fonctionnelle, Théorie et Applications*. Masson, Paris, 1983. English translation : Springer Verlag.
- [DGM98] P. Dupuis, U. Grenander, and M. I. Miller. Variational problems on flows of diffeomorphisms for image matching. *Quart. Appl. Math.*, 56:587–600, 1998.
- [GM98] U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quart. Appl. Math.*, 56:617–694, 1998.
- [HTF09] T. Hastie, R. Tibshirani, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics, 2009.
- [KBS⁺09] Marius Kloft, Ulf Brefeld, Soeren Sonnenburg, Pavel Laskov, Klaus-Robert Müller, and Alexander Zien. Efficient and Accurate Lp-Norm Multiple Kernel Learning. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 997–1005. 2009.
- [RVHR10] L. Risser, R. Vialard, F.X. and Wolz, D.D. Holm, and D. Rueckert. Simultaneous fine and coarse diffeomorphic registration: Application to the atrophy measurement in alzheimer’s disease. In *MICCAI, LNCS*, 2010.
- [RVW⁺11] L. Risser, F. X. Vialard, R. Wolz, D. D. Murgasova, M. and Holm, and D. Rueckert. Simultaneous multiscale registration using large deformation diffeomorphic metric mapping. *IEEE Trans. Image Proc.*, To appear, 2011.

- [Sai88] S. Saitoh. *Theory of reproducing kernels and its applications*. Pitman Research Notes in Mathematics, 1988.
- [Sch64] Laurent Schwartz. Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *Journal d'Analyse Mathématique*, 13:115–256, 1964. 10.1007/BF02786620.
- [SFP⁺10] Nikhil Singh, P. Thomas Fletcher, J. Samuel Preston, Linh Ha, Richard King, J. Stephen Marron, Michael Wiener, and Sarang Joshi. Multivariate statistical analysis of deformation momenta relating anatomical shape to neuropsychological measures. In Tianzi Jiang, Nassir Navab, Josien P.W. Pluim, and Max A. Viergever, editors, *MICCAI*, volume 6363 of *LNCS*, pages 529–537. Springer, 2010.
- [SLNP11] S. Sommer, F. Lauze, M. Nielsen, and X. Pennec. Kernel bundle epdiff: Evolution equations for multi-scale diffeomorphic image registration. In *Scale Space and Variational Methods in Computer Vision*, Lecture Notes in Computer Science. Springer, 2011.
- [SNLP11] S. Sommer, M. Nielsen, F. Lauze, and X. Pennec. A multi-scale kernel bundle for lddmm: Towards sparse deformation description across space and scales. In *Proceedings of IPMI 2011*, Lecture Notes in Computer Science. Springer, 2011.
- [Tro98] A. Trouvé. Diffeomorphic groups and pattern matching in image analysis. *Int. J. Comput. Vision*, 28:213–221, 1998.
- [VMTY04] M. Vaillant, M. I. Miller, A. Trouvé, and L. Younes. Statistics on diffeomorphisms via tangent space representations. *Neuroimage*, 23(S1):S161–S169, 2004.